

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平4-218887

(43) 公開日 平成4年(1992)8月10日

(51) Int.Cl. <sup>5</sup>	識別記号	庁内整理番号	F I	技術表示箇所
G 0 6 K 9/20	3 4 0 P	9073-5L		
G 0 6 F 15/20	5 3 4 Z	6945-5L		
G 0 6 K 9/03		Z 7737-5L		

審査請求 未請求 請求項の数3(全7頁)

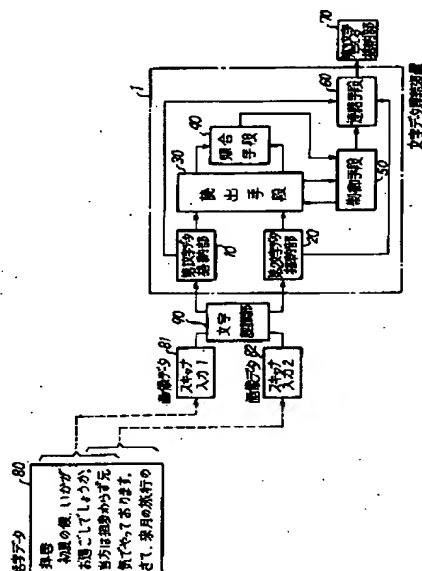
(21) 出願番号	特願平3-69941	(71) 出願人	000002369 セイコーエプソン株式会社 東京都新宿区西新宿2丁目4番1号
(22) 出願日	平成3年(1991)4月2日	(72) 発明者	中林 薫 長野県諏訪市大和3丁目3番5号 セイコーエプソン株式会社内
(31) 優先権主張番号	特願平2-167384	(74) 代理人	弁理士 長谷川 芳樹 (外3名)
(32) 優先日	平2(1990)6月26日		
(33) 優先権主張国	日本(J P)		

(54) 【発明の名称】 文字データ連結装置

(57) 【要約】

【目的】 第1文字データ群と第2文字データ群とが部分的に重複している場合に、自動的に一方の重複した文字データを削除して、2つの文字データを連結させる文字データ連結装置を提供することを目的とする。

【構成】 検出手段によって、第1文字データ格納部に格納された第1文字データ群と、第2文字データ格納部に格納された第2文字データ群との重複部分が検出される。そして、連結手段では、この重複部分の一方が取り除かれ、第1文字データ群と第2文字データ群とが1つの文字データとして連結される。



(2)

特開平4-218887

1

## 【特許請求の範囲】

【請求項1】 任意の部分データ単位ごとにデータアクセスできる第1および第2文字データ格納部と、前記第1文字データ格納部に格納された第1文字データ群と前記第2文字データ格納部に格納された第2文字データ群とを比較して重複する部分データを検出する検出手段と、前記検出手段で検出された重複する部分データの一方を取り除いて、第1文字データ群と第2文字データ群とを連結する連結手段とを備えることを特徴とする文字データ連結装置。

【請求項2】 任意の部分データ単位ごとにデータアクセスできる第1および第2文字データ格納部と、前記第1文字データ格納部に格納された第1文字データ群の行方向または列方向にみた末尾の部分データと前記第2文字データ格納部に格納された第2文字データ群の任意の部分データとを読み出し、この任意の部分データが第2文字データ群の末尾の部分データである場合にこの部分データの読み出し後に終了信号を発する読出手段と、前記読出手段で読み出された第1文字データ群の末尾の部分データと第2文字データ群の任意の部分データの内容を照合する照合手段と、前記照合手段の結果が一致の場合には第2文字データ群の一致した部分データの次の部分データを伝える信号を発し、前記照合手段の結果が不一致の場合で読出手段から終了信号を受けていない場合には第2文字データ群の次の部分データを任意の部分データとして読み出すように前記読出手段に指令を送り、前記照合手段の結果が不一致の場合で前記読出手段から終了信号を受けとった場合には第2文字データ群の先頭の部分データを伝える信号を発する制御手段と、第1文字データ群全体を前記第1文字データ格納装置から読み出し、前記制御手段からの信号によって伝えられた部分データから末尾の部分データまでを前記第2文字データ格納装置から読み出して、第1文字データ群の後に第2文字データ群を連結する連結手段とを備えることを特徴とする文字データ連結装置。

【請求項3】 任意の部分データ単位ごとにデータアクセスできる第1および第2文字データ格納部と、前記第1文字データ格納部に格納された第1文字データ群の行方向または列方向にみた末尾の1または2以上の部分データを読み出し、この部分データと同一個数の部分データを前記第2文字データ格納部に格納された第2文字データ群の先頭から読み出し、第2文字データ群の末尾の部分データまで読み出した場合にこの読み出し後に終了信号を発する読出手段と、前記読出手段で読み出された第1文字データ群と第2文字データ群の1または2以上の部分データの内容を照合する照合手段と、前記照合手段の結果が一致の場合には第2文字データ群から読み出した末尾の部分データの次の部分データを伝える信号を発し、前記照合手段の結果が不一致の場合で前記読出手段から終了信号を受けていない場合には読み出す部分デー

2

タの数を1つ増やすように前記読出手段に指令を送り、前記照合手段の結果が不一致の場合で前記読出手段から終了信号を受けとった場合には第2文字データ群の先頭の部分データを伝える信号を発する制御手段と、第1文字データ群全体を前記第1文字データ格納装置から読み出し、前記制御手段からの信号によって伝えられた部分データから末尾の部分データまでを前記第2文字データ格納装置から読み出して、第1文字データ群の後に第2文字データ群を連結する連結手段とを備えることを特徴とする文字データ連結装置。

## 【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、文字データ連結装置に関するものである。

【0002】

【従来の技術】 文字データの連結は単純に第1文字データ群の末尾と第2文字データ群の先頭を連結していた。

【0003】

【発明が解決しようとする課題】 しかし、単純な連結では不便な場合がある。例えば、雑誌の記事などの長い文章をハンディスキャナを用いて複数回に分けて読み取って、この画像データを文字認識手段で文字データとした場合、1回の読み込みで得られる文字データと次の読み込みで得られる文字データの一部が重複してしまうことがある。このような場合に、文字データを単純に連結すると重複部分を後から人手で確認して削除しなければならなかった。

【0004】 本発明の目的は、第1文字データ群と第2文字データ群の一部が重複していても、自動的にその部分を探索して、一方の重複部分を取り除いて2つの文字データ群を連結する文字データ連結装置を提供しようというものである。

【0005】

【課題を解決するための手段】 第1の発明の文字データ連結装置には、任意の部分データ単位ごとにデータアクセスできる第1および第2文字データ格納部と、第1文字データ格納部に格納された第1文字データ群と第2文字データ格納部に格納された第2文字データ群とを比較して重複する部分データを検出する検出手段と、検出手段で検出された重複する部分データの一方を取り除いて、第1文字データ群と第2文字データ群とを連結する連結手段とが備えられている。

【0006】 また、第2の発明の文字データ連結装置には、任意の部分データ単位ごとにデータアクセスできる第1および第2文字データ格納部と、第1文字データ格納部に格納された第1文字データ群の行方向または列方向にみた末尾の部分データと第2文字データ格納部に格納された第2文字データ群の任意の部分データとを読み出し、この任意の部分データが第2文字データ群の末尾の部分データである場合にこの部分データの読み出し後

3

に終了信号を発する読出手段と、読出手段で読み出された第1文字データ群の末尾の部分データと第2文字データ群の任意の部分データの内容を照合する照合手段と、照合手段の結果が一致の場合には第2文字データ群の一致した部分データの次の部分データを伝える信号を発し、照合手段の結果が不一致の場合で読出手段から終了信号を受けていない場合には第2文字データ群の次の部分データを任意の部分データとして読み出すように読出手段に指令を送り、照合手段の結果が不一致の場合で読出手段から終了信号を受けとった場合には第2文字データ群の先頭の部分データを伝える信号を発する制御手段と、第1文字データ群全体を第1文字データ格納装置から読み出し、制御手段からの信号によって伝えられた部分データから末尾の部分データまでを第2文字データ格納装置から読み出して、第1文字データ群の後に第2文字データ群を連結する連結手段とが備えられている。

【0007】さらに、第3の発明の文字データ連結装置には、任意の部分データ単位ごとにデータアクセスできる第1および第2文字データ格納部と、第1文字データ格納部に格納された第1文字データ群の行方向または列方向にみた末尾の1または2以上の部分データを読み出し、この部分データと同一個数の部分データを第2文字データ格納部に格納された第2文字データ群の先頭から読み出し、第2文字データ群の末尾の部分データまで読み出した場合にこの読み出し後に終了信号を発する読出手段と、読出手段で読み出された第1文字データ群と第2文字データ群の1または2以上の部分データの内容を照合する照合手段と、照合手段の結果が一致の場合には第2文字データ群から読み出した末尾の部分データの次の部分データを伝える信号を発し、照合手段の結果が不一致の場合で読出手段から終了信号を受けていない場合には読み出す部分データの数を1つ増やすように読出手段に指令を送り、照合手段の結果が不一致の場合で読出手段から終了信号を受けとった場合には第2文字データ群の先頭の部分データを伝える信号を発する制御手段と、第1文字データ群全体を第1文字データ格納装置から読み出し、制御手段からの信号によって伝えられた部分データから末尾の部分データまでを第2文字データ格納装置から読み出して、第1文字データ群の後に第2文字データ群を連結する連結手段とが備えられている。

【0008】

【作用】第1の発明の文字データ連結装置によれば、読出手段によって、第1文字データ格納部に格納された第1文字データ群と、第2文字データ格納部に格納された第2文字データ群との重複部分が検出される。そして、連結手段では、この重複部分の一方が取り除かれ、第1文字データ群と第2文字データ群とが1つの文字データとして連結される。

【0009】第2の発明の文字データ連結装置によれば、まず、読出手段によって、第1文字データ格納部に

4

格納された第1文字データ群の行方向または列方向にみた末尾の部分データが読み出される。また、第2文字データ格納部に格納された第2文字データ群の任意の部分データが読み出される。次に、照合手段によって、第1文字データ群の末尾の部分データと第2文字データ群の任意の部分データの内容が照合される。この照合手段の照合の結果が一致の場合には、第2文字データ群の一致した部分データの次の部分データを伝える信号が発せられる。また、この照合手段の結果が不一致の場合で読出手段から終了信号を受けていない場合には、第2文字データ群の次の部分データを任意の部分データとして読み出すように読出手段に指令が送られる。さらに、照合手段の結果が不一致の場合で読出手段から終了信号を受けとった場合には、第2文字データ群の先頭の部分データを伝える信号が発せられる。そして、連結手段では、第1文字データ群の全部が第1文字データ格納装置から読み出され、制御手段からの信号によって伝えられた部分データから末尾の部分データまでが第2文字データ格納装置から読み出され、第1文字データ群の後に第2文字データ群が連結される。

【0010】第3の発明の文字データ連結装置は、1または2以上の部分データを照合して重複部分を検出する点が第2の発明の文字データ連結装置と異なる。つまり、第3の発明の文字データ連結装置の読出手段では、第1文字データ群の行方向または列方向にみた末尾の1または2以上の部分データが読み出され、また、この部分データと同一個数の部分データを第2文字データ群の先頭から読み出される。そして、これらの部分データが照合手段で照合されて、重複部分が検出されるのである。

【0011】

【実施例】以下、第1の発明から第3の発明までの一実施例について、図1から図6を用いて説明する。図1は本実施例の構成ブロック図である。本実施例の文字データ連結装置1は、第1文字データ群が格納された第1文字データ格納部10と、第2文字データ群が格納された第2文字データ格納部20と、第1文字データ格納部10と第2文字データ格納部20から行または列または文字列の部分データを読み出す読出手段30と、読出手段30で読み出された部分データを照合する照合手段40と、読出手段30を制御する制御手段50と、第1文字データ群と第2文字データ群を連結する連結手段60とから構成されている。第1文字データ格納部10および第2文字データ格納部20は、任意の部分データ単位ごとにデータアクセスできるデータ構造を備えている。この任意の部分データには、例えば、行単位の部分データや、列単位の部分データや、文節単位の部分データや、1文字単位の部分データなどがある。また、連結手段で接続された文字データは文字データ3格納部70に格納される。

【0012】次に、本実施例の処理の内容について、図2のフローチャートを用いて説明する。まず、手紙などの活字データ80の上半分をハンディスキャナで読み込み、画像データ81を獲得する(ステップ100)。この画像データ81を文字認識部90で画像解析して、第1文字データ群120を生成する(ステップ101)。また、活字データ80の下半分をハンディスキャナで読み込み、画像データ82を獲得する(ステップ102)。この画像データ82を文字認識部90で画像解析して、第2文字データ群121を生成する(ステップ103)。そして、第1文字データ120群を第1文字データ格納部10に、第2文字データ群121を第2文字データ格納部20にそれぞれ格納する。このように格納された第1文字データ群120の3、4行目と第2文字データ群121の1、2行目は重複しているもので、以後の処理でこの内の一方を取り除くのである。

【0013】重複した行を取り除くために、まず、第1文字データ格納部10と第2文字データ格納部20から部分データである1行ごとの文字データを読み出す手段30を用いて読み出す。具体的には、まず、第2文字データ群121の行数を示す検索カウンタに「1」(1行目を示す。)を設定する(ステップ104)。次に、第1文字データ群120の末尾の行を読み出す(ステップ105)。そして、ステップ104で設定された検索カウンタを用いて、第2文字データ群121の任意の1行を読み出し(ステップ106)、読み出し後に検索カウンタに1を加える(ステップ107)。ステップ106からステップ111まではループ構造になっており、検索カウンタによって第2文字データ群121の第1行目から順番に文字データを読み出していく。そして、この処理で第2文字データ群121の末尾の行まで読み出したために、ステップ106で正しく文字データを読み出せなくなった場合(ステップ108)には、検索カウンタ「1」にした上で(ステップ109)、連結手段の処理を行う。また、末尾の行まで読み出していない場合には、照合手段40に処理を移し、第1文字データ群120と第2文字データ群121からそれぞれ読み出された文字データを照合する(ステップ110)。この照合によって、文字データが一致した場合には、連結手段60に処理を移す(ステップ111)。また、一致しない場合には、ステップ106に処理を戻して、第2文字データ群121の次の行の文字データを読み出す。連結手段60では、第1文字データ群の全部の行と第2文字データ群の検索カウンタで示す行以降を連結する(ステップ112)。

【0014】なお、照合手段の処理をより正確に行うためには、複数行を用いて比較処理を行えばよい。この場合には、まず、ステップ105で第1文字データ群の末尾の行から順番に検索カウンタで示す行数だけ読み出す。また、ステップ106でも同様に、第2文字データ

群の1行目から検索カウンタが示す行数だけ読み出す。そして、ステップ110で第1文字データ群と第2文字データ群の複数行の照合を行うのである。

【0015】次に、本実施例で連結される文字データの例を、図3～図5の文字データ連結図を用いて説明する。

【0016】図3(a)は第1文字データ群の3、4行目と第2文字データ群の1、2行目が重複している場合の連結例である。この例では、第1文字データ群の末尾に、第2文字データ群の3行目を接続して、第3文字データを作成している。図3(b)についても同様である。

【0017】図4(a)は第1文字データ群の3～5行目と第2文字データ群の1～3行目が重複している場合の連結例である。この例では、第1文字データ群の末尾に、第2文字データ群の4行目を接続して、第3文字データを作成している。図4(b)は第1文字データ群の17～19行目と第2文字データ群の1～3行目が重複している場合の連結例である。この例では、第1文字データ群の末尾に、第2文字データ群の4行目を接続して、第3文字データを作成している。

【0018】図5(a)は第1文字データ群の文字列「初夏の候」が第2文字データ群に存在しない場合の連結例である。この例では、第1文字データ群の末尾に、第2文字データ群の先頭を接続して、第3文字データを作成している。第5図(b)は第1文字データ群の6ワード目の文字列「Epson」が第2文字データ群の4ワード目と重複している場合の連結例である。この例では、第1文字データ群の末尾に、第2文字データ群の4ワード目を接続して、第3文字データを作成している。

【0019】次に、本実施例をパーソナルコンピュータで実現した場合のハードウェア構成について、図6を用いて説明する。パーソナルコンピュータ210は、制御手段50として動作するCPU211と、読出手段30や照合手段40などのプログラムが格納されたRAM212と、BIOSなどのシステムプログラムが格納されたRAM213とから構成される。また、このパーソナルコンピュータ210には、第1文字データ群や第2文字データ群などを表示するCRT220と、ユーザからの指令が入力されるキーボード221と、雑誌の記事などを入力するハンディスキャナ222と、ハンディスキャナ222などで入力された文字データが格納される磁気ディスク223とが備えられている。

【0020】本実施例は、特にハンディスキャナ222を用いて入力した複数の文字データを連結する際に効果を発揮する。つまり、ハンディスキャナ222は入力部の幅が小さいため、通常は数回に分けて文字データ上を走査させる必要がある。この走査においては、入力抜けが生じないように端の部分を重複させながら入力させることが一般に行われている。本実施例を用いれば、この

(5)

特開平4-218887

7

時に生じる重複部分の削除が自動的に行えるのである。

【0021】なお、本実施例で行または列または文字列のいずれの部分データを用いるかは、ユーザによる指令によって決定しても良いし、あるいは2以上の部分データを併用しても良い。併用する例としては、まず行単位で重複部分の照合を行い、次に列単位で重複部分の照合を行って、連結する手段などがある。

【0022】

【発明の効果】本発明の文字データ連結装置であれば、複数の文字データの一部が重複する場合に、自動的に重複部分を探し出して、一方の重複部分を取り除いて文字データを連結することができる。

【0023】したがって、複数の文字データを連結する作業時に、利用者が重複部分を手作業で削除する必要がなくなる。このために、修正ミスの発生を防ぐことができ、さらに手間と時間が節約できる。

【0024】特に、本発明の文字データ連結装置の応用として、雑誌の記事などの長い文章をハンディスキャナで複数回に分けて読み取り、この画像データを文字認識手段で文字データとし、そして、このようにして得られた複数の文字データを連結させて元の文章を再現する

8

場合に効果を発揮する。

【図面の簡単な説明】

【図1】本実施例の構成ブロック図である。

【図2】本実施例の処理の内容を示すフローチャートである。

【図3】本実施例の文字データの連結処理を示す概念図である。

【図4】本実施例の文字データの連結処理を示す概念図である。

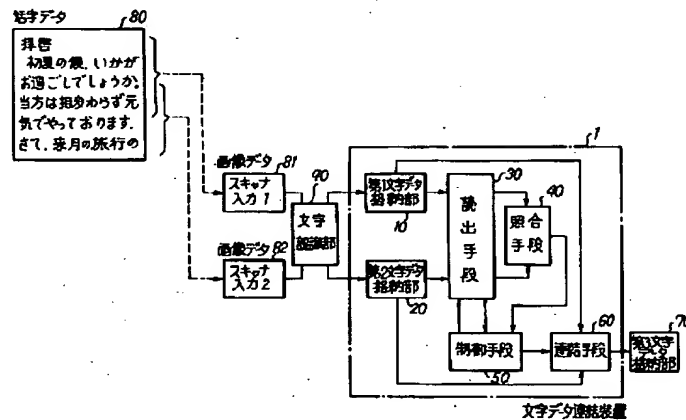
【図5】本実施例の文字データの連結処理を示す概念図である。

【図6】本実施例のハードウェア構成図である。

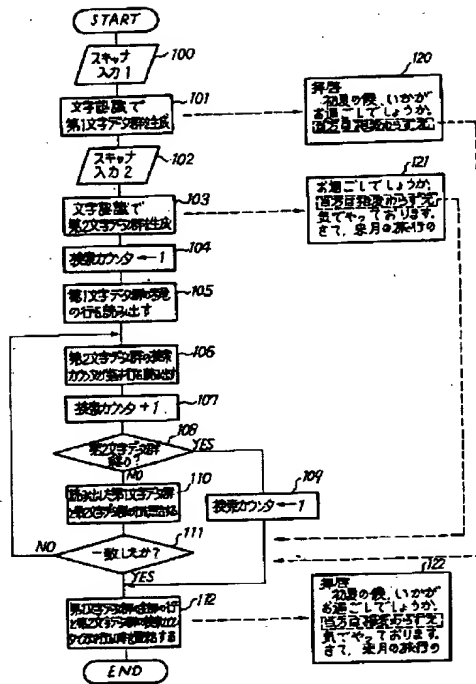
【符号の説明】

- 1…文字データ連結装置
- 10…第1文字データ格納部
- 20…第2文字データ格納部
- 30…読出手段
- 40…照合手段
- 50…制御手段
- 60…連結手段

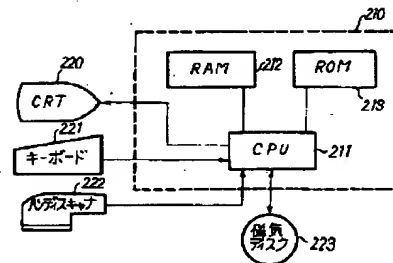
【図1】



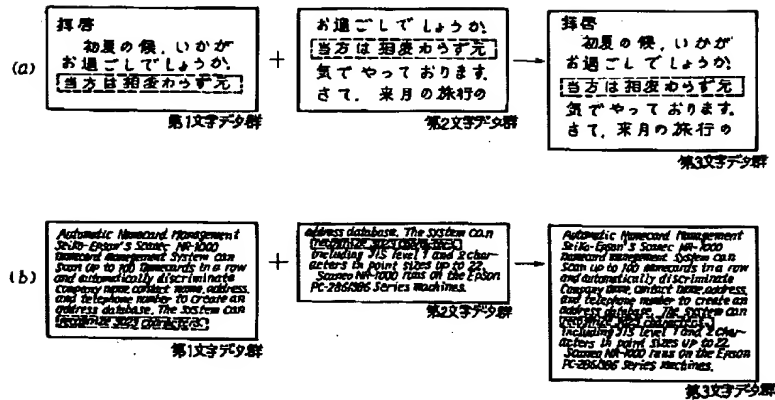
【図2】



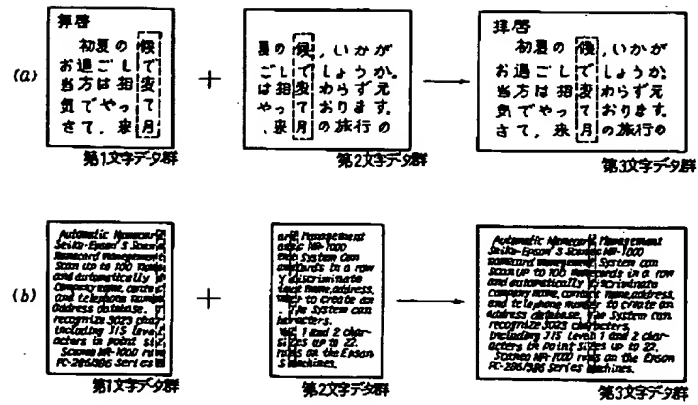
【図6】



【図3】



【図 4】



【図 5】

